

John von Neumann Institute for Computing



Engrailed Homeodomain Folds Overnight by 100 Processors

K. V. Klenin, W. Wenzel

published in

*From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. 40, ISBN 978-3-9810843-6-8, pp. 249-252, 2008.*

© 2008 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume40>

Engrailed Homeodomain Folds Overnight by 100 Processors

Konstantin V. Klenin^{1,2} and Wolfgang Wenzel^{1,2}

¹ Institut für Nanotechnologie, Forschungszentrum Karlsruhe,
Postfach 3640, 76021 Karlsruhe, Germany

² DFG-Centrum für Funktionelle Nanostrukturen, Universität Karlsruhe,
Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany
E-mail: {klenin, wenzel}@int.fzk.de

It has been shown that the native structure of a small protein can be efficiently found as the global minimum of a certain all-atom forcefield. In the present study, we use this approach to simulate folding of Engrailed homeodomain (PDB code IENH) containing the helix-turn-helix motif. The search procedure is based on the Monte Carlo simulated annealing combined with the evolutionary algorithm. We propose a new way of increasing the efficiency of this method.

1 Introduction

The free-energy approach has delivered promising results for protein folding and structure prediction in recent years. Following Anfinsen's hypothesis¹, the native state is postulated to be the global minimum of a all-atom free-energy function. This minimum can be found by optimization methods² involving the Monte Carlo simulated annealing. This approach was successfully used to fold α -, β - and mixed proteins of small length³⁻⁵.

This procedure requires an accurate, transferable energy function, such as the Protein Force Field (PFF02) developed in our group^{6,7}:

$$E = E_{LJ} + E_{ES} + E_{SASA} + E_{HB} + E_R.$$

It contains the following terms: (1) E_{LJ} , the standard Lennard-Jones potential, (2) E_{ES} , the Coulomb energy of electrostatic interaction with efficient dielectric constants, (3) E_{SASA} , a term proportional to the solvent accessible surface area, (4) E_{HB} , a term for the hydrogen bonding, (5) E_R , a term stabilizing the β -regions in the Ramachandran plots. All the atoms are explicitly represented (only the apolar group CH_n is considered as a single atom). The bond angles and the bond lengths are fixed. The degrees of freedom considered are the backbone (ψ, ϕ) and the sidechain (χ_i) dihedral angles. The solvent is taken into account implicitly.

The Monte Carlo procedure involves two kinds of moves: (1) unbiased, consisting of a random change of the dihedral angles, and (2) biased, which set the dihedral angles (within a single residue) to predefined values from a certain library.

The most promising optimization method is the evolutionary algorithm⁸, where a fixed size population of conformations evolves simultaneously. One cycle of the algorithm consists of the following steps. An individual conformation is taken randomly from the population and is subjected to the Monte Carlo simulated annealing starting from a random temperature. The obtained structure is then added to the population. If the similar (in terms of RMSD) conformations are already present, than the only one among them with

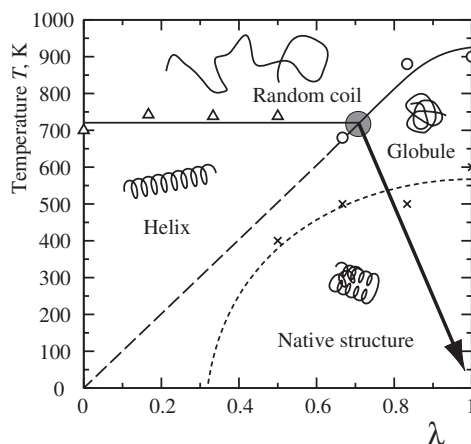


Figure 1. The phase diagram for the Engrailed homeodomain. The optimal annealing path is shown by the arrow.

the lowest energy is kept in the population and the others are removed. Otherwise, the highest energy structure within the entire population is removed (provided the population size has exceeded the certain limit). After many such cycles, the lowest energy structure approaches the global minimum.

In the present work, we improve this procedure by optimizing the conditions of the simulated annealing. The efficiency of our approach is demonstrated by the example of the Engrailed homeodomain (PDB code 1ENH), a small three-helix protein containing the helix-turn-helix motive that is involved in the DNA binding⁹.

2 Methods

It has been observed that during the simulated annealing, as the temperature lowers down, the collapse of the chain occurs first, and only then the secondary structure forms. But in the collapsed conformation the Monte Carlo moves are essentially less efficient. To address this problem, we used the modified energy

$$E = \lambda E_{\text{SASA}} + E_{\text{HB}} + \dots,$$

where λ is an arbitrary parameter in the range $0 \leq \lambda \leq 1$. Consider the (artificial) phase diagram in the coordinates λ and the temperature T as shown in Fig. 1. (By the “phase” we imply the most probable state of the system.) The “triple” point is characterized by the most extensive fluctuations of the structure in terms of its size and energy. The main idea of this work is to use it as the starting point for the annealing process. (The unrealistic high temperature is due to the stabilizing effect of the biased moves.) First, the system is kept at the fixed λ and T for a period of approximately one relaxation time τ and then, for the same time period, it is moved along the arrow in Fig. 1.

The triangle and circle points in Fig. 1 were obtained without knowledge of the native structure as illustrated in Fig. 2.

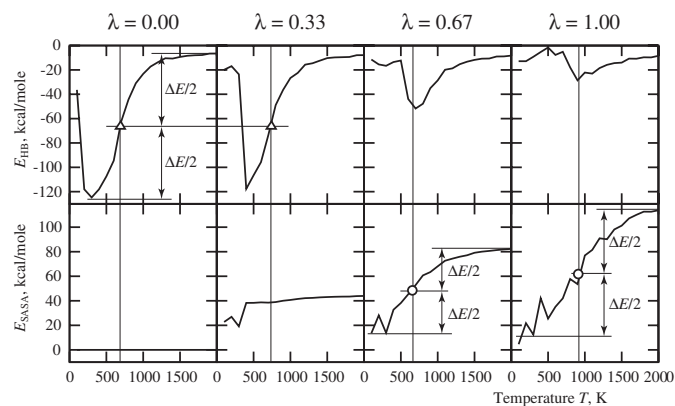


Figure 2. The mean energies E_{HB} and E_{SASA} (for the quasi-stable states reached by the Monte Carlo process starting from an extended conformation) as functions of the temperature T for various λ . The boundary of the helical phase corresponds to the middle value of E_{HB} at $\lambda = 0$. The boundary of the globular phase corresponds to the middle value of E_{SASA} at each particular λ .

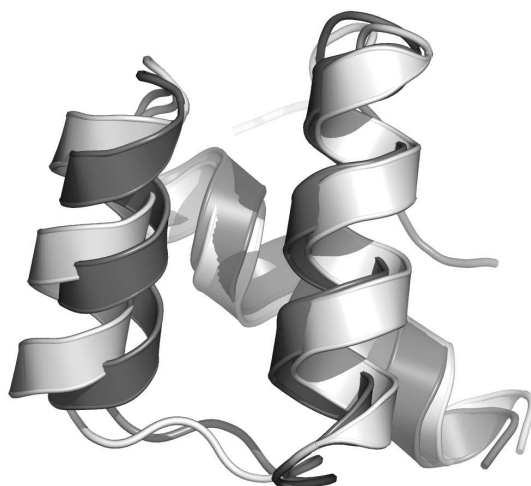


Figure 3. Comparison between the simulated (dark gray) and experimental (light gray) structures. The backbone RMSD is 2.8 Å.

3 Results and Conclusions

The comparison between the simulated and experimental structures is shown in Fig. 3. When the “triple” point of the phase diagram is known, the native structure of the Engrailed homeodomain can be found in 2 months CPU time. Using the optimized annealing path increases the efficiency of the evolutionary algorithm by an order of magnitude.

Acknowledgments

We thank the BMBF, the Fond der Chemischen Industrie, the Deutsche Forschungsgemeinschaft (grants WE 1863/10-1, WE 1863/10-2, WE 1863/14-1) and the Kurt Eberhard Bode Stiftung for financial support. We are thankful to the Barcelona Supercomputer Center and to the Korea Institute of Science and Technology for computational resources.

References

1. C. B. Anfinsen, *Principles that govern the Folding of Protein Chains*, Science **181**, 223, 1973.
2. A. Schug, A. Verma, W. Wenzel, and G. Schoen, *Biomolecular structure prediction with stochastic optimization methods*, Adv. Eng. Materials **7**, 1005, 2005.
3. A. Verma, S. Murthy, K. H. Lee, E. Starikov, and W. Wenzel, *De novo all atom folding of helical proteins*, NIC Series **34**, 45, 2006.
4. W. Wenzel, *Predictive folding of a β hairpin in an all-atom free-energy model*, Europhys. Letters **76**, 156, 2006.
5. S. M. Gopal and W. Wenzel, *De Novo Folding of the DNA-Binding ATF-2 Zinc Finger Motif in an All-Atom Free-Energy Forcefield*, Angew. Chemie Int. **45**, 7726, 2006.
6. T. Herges and W. Wenzel, *An All-Atom Force Field for Tertiary Structure Prediction of Helical Proteins*, Biophysical Journal **87**, 3100, 2004.
7. A. Verma and W. Wenzel, *Towards a universal Free Energy Forcefield for All atom Protein Folding*, (Submitted) 2008.
8. A. Verma, S. M. Gopal, J.S. Ooh, K.H. Lee, and W. Wenzel, *All atom de-novo protein folding with a scalable evolutionary algorithm*, J. Comput. Chem. **28**, 2552, 2007.
9. N.D. Clarke, C.R. Kissinger, J. Desjarlais, G.L. Gilliland, C.O. Pabo, *Structural studies of the engrailed homeodomain*, Protein Sci. **3**, 1779, 1994.